Psychological Monographs

General and A[illegible]d

No. 448
1957

Fruchter, Blake, and Mouton

Vol. 71
No. 19

Some Dimensions of Interpersonal Relations in Three-Man Airplane Crews

By

Benjamin Fruchter, Robert R. Blake, Jane Srygley Mouton

University of Texas

Price $1.00

Edited by Herbert S. Conrad

Published by The American Psychological Asso[illegible]

Psychological Monographs: General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

MANUSCRIPTS should be sent to the Editor.

Because of lack of space, the *Psychological Monographs* can print only the original or advanced contribution of the author. *Background and bibliographic materials must, in general, be totally excluded* or kept to an irreducible minimum. Statistical tables should be used to present only the most important of the statistical data or evidence.

Preparation of manuscripts for publication as monographs should follow the procedures given in the APA *Publication Manual*, 1957 Revision—this revision supersedes all previous manuals for preparing manuscripts for publication as monographs.

CORRESPONDENCE CONCERNING BUSINESS MATTERS (such as subscriptions and sales, change of address, etc.) should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee third-class forwarding postage.

Vol. 71, No. 19 Whole No. 448, 1957

Psychological Monographs: General and Applied

Some Dimensions of Interpersonal Relations in Three-Man Airplane Crews

BENJAMIN FRUCHTER, ROBERT R. BLAKE, AND JANE SRYGLEY MOUTON
University of Texas[1]

CONSTRUCTION OF THE CREW INTERACTION SCALES

Introduction

ONE of the interesting characteristics of a number of modern technological devices and systems is that their operation requires the team work of several people. This consideration leads to questions concerning the extent to which the interpersonal relations, which are part of the basis of organizational efficiency, can be measured and predicted.

The purposes of this study were: (*a*) to construct a scale for assessing the interpersonal relations of three-man, B-47, medium-bomber crews consisting of aircraft commander (AC), observer (O), and copilot (P); (*b*) to determine the basic content or factors represented in the items; and (*c*) to estimate the reliability and validity of the resulting homogeneous scales.

Rationale for Selecting Items

Several conceptual approaches to group phenomena were considered in selecting items to be included in the initial form of the instrument.

Theory and experiments in the field of group dynamics (5) which have led to the development of concepts such as "group standard" and "group cohesiveness" were reviewed, and items suitable for measuring these dimensions in B-47 crews were developed. The approach by Homans (3)—which relates the group work system to its social system through variables such as liking and amount of interaction—also was examined, with items being written to measure these aspects of group behavior. Recent developments in group therapy (7) that emphasize group and individual relations between members and the leader, and also among members, both from a choice and perceptual standpoint, and that have resulted in experiments on the accuracy of interpersonal perceptions in small groups were examined; and items suggested by developments in this area also were included. Scales that had been developed for measuring the relationships among B-29 crew members were examined; some items from such scales were included because of their demonstrated value in measuring interpersonal relations. Finally, the sociometric literature was surveyed for additional items that had been found useful for evaluating interpersonal relations in a variety of work situations.

Items from each source were adapted for measuring relations among B-17 crew members. A final scale was developed after a tentative scale had been devised and evaluated through interviews with B-47 crew members and all items had been found acceptable in the sense that they assess relationships that crew members considered to be relevant in evaluating their crew performance. On the basis of nontechnical considerations concerning unsuitability of items for use in measuring crew relations, pointed out by flying and operational personnel, many items which otherwise appeared useful for scaling interpersonal relationships were deleted.

Description of the Crew Interaction Scale

The Crew Interaction Scale consisted of 44 items concerning a variety of aspects of crew relations.

The *first 22 items* required each crew member to rank himself and each of the other two members on some aspect of each member's crew

[1] This research was supported in part by the USAF under Contract Number AF 18(600)-602, monitored by the 3305th Research and Development Group (Combat Crew Training Research Laboratory), Human Resources Research Center, Randolph Air Force Base, Texas.

performance. Items included the extent to which each dominates the decision making of the crew, accepts responsibility for both air and ground activities, is agreeable and willing to go along with crew decisions, and exhibits self-control during periods of stress in the air.

The *next five items* dealt with how each crew member relates to the others. They included items such as the extent to which each crew member sees himself as clashing with the others, the degree to which each member feels he can influence the others to change decisions, and how much each feels a handicap in doing his job because of not obtaining necessary information and instructions from other members.

The *final 17 items* called for evaluation by its members of the crew as a whole. Items such as the degree to which the crew functions on the basis of an informal code of procedure and the extent to which members see the crew as a well organized team were included in this section.

The Ranking-Rating Technique

In a preliminary laboratory investigation with three-man groups, rankings had been found to be satisfactorily consistent from group to group (1). In addition to using the ranking method, it was also considered desirable to devise a rating scale for determining the *relative strength* of each ranking. The crew members were given the following directions:

> In filling out the following items think of the three members of your crew—yourself and the two others—and try to characterize how each behaves in terms of the scale provided. Use the symbol AC for Aircraft Commander, P for Pilot, and O for Observer to identify yourself and the other two crew members. *Make an effort to put the identification symbol of only one crew member in any one box.*

Depending on the characteristics judged, two types of scales were used. One is a bipolar scale with a neutral point in the middle, and the other is an all-to-none scale with a 50-per-cent point at the center. In constructing a scale, the intervals were kept as uniform as possible by using the same or similar modifying phrases to define positions along the continuum. The following is an example of the type of scale used to measure bipolar characteristics:

> Rank the members of your crew for the extent to which they accept or reject responsibility in both air and ground activities.
>
> ☐ always rejects responsibility ☐ extremely often ☐ very often
>
> ☐ often ☐ neither accepts nor rejects responsibility ☐ often accepts responsibility
>
> ☐ very often ☐ extremely often ☐ always accepts responsibility

An example of a scale used to measure characteristics that can be scaled along an all-to-none continuum is:

> Rank the members of your crew for the frequency with which they make errors in the performance of crew activities in the air.
>
> ☐ always makes errors ☐ extremely often ☐ very often
>
> ☐ often ☐ makes errors some of the time ☐ a little
>
> ☐ very little ☐ almost never ☐ never makes errors

According to the instructions for both kinds of scales, each crew member placed his rankings of the three crew members (including himself) at the appropriate positions on the nine-point continuum. Tied ratings were not allowed. Some crew members objected to being forced occasionally to make a discrimination, as was required by the "no ties" restriction. The objection was met by allowing the rater to place more than one person within a scale interval and indicating the rank-ordering within the interval.

Items Included in the Scale

The following are the 44 items on which each crew member rated *himself and the other two members of his crew.*

1. Rank the members of your crew for the extent to which they accept or reject responsibility in both air and ground activities.

2. Rank the members of your crew in terms of being agreeable and willing to go along with crew decisions in both air and ground activities.

3. Rank the members of your crew in terms of the extent they take the initiative in getting things done in both air and ground activities.

4. Rank the members of your crew for the frequency with which they make errors in the performance of crew activities in the air.

5. Rank the members of your crew for the extent to which they work cooperatively with the crew in both air and ground activities.

6. Rank the members of your crew in terms of the amount of leadership they provide during air activities.

7. Rank the members of your crew for the extent to which they keep the crew "on the ball" in both air and ground activities.

8. Rank the members of your crew for the extent to which their actions are well-timed and appropriate to the actions of the other members of the crew in air activities.

9. Rank the members of your crew for the extent to which they place crew welfare above personal considerations in both air and ground activities.

10. To what extent do the members of your crew follow the directions and suggestions of the others in critical air situations?

11. Rank the members of your crew for the extent to which they put your crew at ease and make being a member of the crew more enjoyable in both air and ground activities.

12. Rank the members of your crew for the extent to which they feel frustrated by being a member of this crew.

13. To what extent do the members of your crew usually agree about what should be done in the air?

14. Rank the members of your crew for their competence, in air activities, in the specialty to which they are assigned.

15. Rank the members of your crew in terms of the extent to which they want to do a good job in both air and ground activities.

16. Rank the members of your crew for self-control during periods of stress in the air.

17. Rank the members of your crew for clarity in giving directions both in air and ground activities.

18. Rank the members of your crew for their satisfaction with the crew's over-all performance in both air and ground activities.

19. Rank the members of your crew for the extent to which they dominate the decision making of the crew in both air and ground activities.

20. Rank the members of your crew in terms of the extent to which they keep reminding the crew of regulations and SOP in both air and ground activities.

21. Rank the members of your crew in terms of the extent to which they are interested in their technical specialty in both air and ground activities.

22. Rank the members of your crew for the extent to which they are consistent (level of performance same from day to day) in their performance of activities in the air.

23. Rank the other members of your crew in terms of the degree you clash with them in crew discussions both in the air and on the ground.

24. Rank the two other members of your crew in terms of your liking for them in both air and ground relationships.

25. Rank the other two members of your crew in terms of the ease with which you can influence them to change their decisions in both air and ground activities.

26. Rank the other two members of your crew for the readiness with which you would expect them to back you up if a misunderstanding arose in your crew in ground activities.

27. Rank the other two crew members in terms of how often you are handicapped in doing your job because he has not given you necessary information or instructions in both ground and air activities.

28. To what extent does your crew work as a well organized team rather than as a collection of individuals in both air and ground activities?

29. To what extent do you feel "left out" of important decisions that your crew makes in both air and ground activities?

30. How well satisfied are you with the way you, personally, have worked out in this crew in its air activities?

31. What is your attitude toward remaining with your present crew?

32. Rate your crew for its enthusiasm for flying in present crew assignments.

33. Rate your crew for the speed with which it comes to agreement when decisions have to be made in the air.

34. Rate your crew in terms of its willingness to fly *under adverse conditions.*

35. Rate your crew for the extent to which it is a well organized team in both air and ground activities.

36. Rate your crew for its confidence in its ability to perform in critical situations in the air.

37. Rate the morale of your crew in both air and ground activities.

38. Rate your crew for the extent to which

it sticks together in the face of disrupting factors in both air and ground activities.

39. Rate your crew for the extent to which it is a "stickler" for following formal or official regulations and SOP in both air and ground activities.

40. Rate your crew for its effectiveness in planning together in both air and ground activities.

41. Rate your crew for the extent to which it is able to maintain a consistent level of flying efficiency.

42. Rate your crew for the frequency with which it makes errors in carrying out its air assignments.

43. Rate your crew for its "cockpit confusion."

44. Rate your crew for the extent to which it functions in terms of an informal code of procedures that it has worked out by itself in its air activities.

Sample

The Crew Interaction Scale was administered to 90 male flying officers, who composed 30 intact B-47 crews in operational training. Crews had been organized for approximately four to five months. All were tested individually for an hour. Data collection took place over a one-week period since only small numbers of personnel could be made available in any one test period.

Item Analysis

Agreement in Crew Ratings by Items

Since each member rated either the crew as a whole or specifically designated members, several ratings of a crew and its members were obtained from each item. From a statistical point of view the means of the several ratings should represent a better estimate of the characteristic being judged than any of the ratings considered separately. To represent useful estimates it is desirable that inter-crew means be differentiated adequately relative to the individual ratings on the characteristic being judged. Horst (4) has developed a procedure which can be applied to estimating the extent of agreement in ratings for any one crew relative to the discrimination between crews. If ratings are internally consistent, in the sense that the several judgments within each crew or by each crew member agree perfectly, the Horst coefficient would be +1.0. As the spread of the ratings around the mean of the within-crew ratings increases, the index approaches zero. Horst has demonstrated that the formula is related to the Spearman-Brown method for estimating reliability and to the Kuder-Richardson formula (no. 21) for internal consistency.

Each rating was weighted from zero at the "most desirable" end of the scale to eight at the "least desirable" end. Agreement indexes were computed for the following combined-score ratings. (*a*) The variance of the sum of the ratings of the crews on a given item was evaluated against the within-crew variance. (*b*) Next, the variance of the sum of the ratings of the crews on a given item was also evaluated against all the within-crew variance *exclusive of the self-ratings*. (*c*) Finally, the variance of the sum of the ratings for a *crew position* on a given item was evaluated against the within-crew variance of those ratings.

The *Horst agreement coefficients* for the scores included in scales are shown in Tables 5 through 10. The results for all of the items may be summarized as follows:

The interrater agreement indexes for the crew totals on the first 27 items (of which the first 22 were based on nine ratings and the next 5 on six ratings each per crew) were satisfactorily high with few exceptions. The results indicated both an acceptable consensus among crew members in rating one another and a satisfactory discrimination between crew means.

There were only three sets of ratings

for each crew on the last 17 items. As computed by the Horst method, these items had a wider range of interrater agreement values. Also, from the standpoint of the kind of judgments involved, items 29, 30, and 31 were different from the others, since the respondent was required to evaluate his own relationship to the crew rather than to rate the crew as a unit. Such items might be suitable for differentiating among individuals, but they would not be suitable for computing mean crew ratings. Indexes of agreement were therefore not obtained for these three items. Mean ratings for items 39 and 44 did not adequately discriminate among crews. Both refer to the extent to which the crew functions in accordance with a prescribed rather than an informal group structure. The agreement values for items 33, 34, and 40, concerned with speed in reaching crew agreement, readiness to fly under adverse conditions, and crew planning effectiveness, were also unsatisfactorily low.

The general conclusions from assessing the consistency of crew ratings by the Horst method were that crew members do agree in their judgments on most items and that there is satisfactory discrimination among the means of different crews. Items were next considered from the standpoint of validity.

Criteria Used for Validation

Two sets of proficiency rankings were procured from supervisory personnel at Lake Charles Air Force Base. They were a ranking of crews on a wing basis and a ranking of crews by squadron, within the wing.

The 67th Wing had been assembled as an operational unit four or five months previously. The Wing Standardization Board, which was composed of skilled personnel representing each of the three specialties, had completed the evaluation of the crews used in the validation study during the week previous to the test administration. In making its evaluations the Wing Standardization Board used a fixed set of criteria, and, on the basis of the evaluation of specific components of individual performance and coordination among crew members, assigned a ranking for each crew in the wing.

Crews within each squadron were ranked in terms of performance efficiency as judged by three squadron commanders, who based their rankings on information supplied both by flight commanders and the squadron operations officers. The two most important considerations in making these ratings were the circular error in navigation and bombing assignments and the level of performance of the Observer.

Rankings by squadron and wing were used to distribute the crews into high, middle, and low performance groups to be used in validating the Crew Interaction Scale. The three performance groups were determined by combining the standings given by the Wing Standardization Board with squadron rankings. Each of the three sets of rankings was given an equal weight and the rankings summed into a single score. The 10 crews with the highest combined rankings were identified as the high group, the next 10 as the middle group, and the lowest 10 as the low performance group.

Item Validity

The validity for each item was determined in the following manner. An integral score value was located as close as possible to the median of the distribution for each item. The number of item-ratings above and below the median score for each of the three criterion groups was then determined. The following diagram shows how the data were set up for an item:

		Performance		
		Low	Middle	High
Median	Above	2	5	7
	Below	8	5	3

(Item 1 AC_i)

A chi-square test for the significance of the differences between the number above and below the median in the three criterion groups was made, and the contingency correlation coefficient, corrected for coarse grouping, was calculated for those items having significant chi squares.

The results indicated that many items are related significantly to the criterion groups at, or below, the 20-per-cent level of confidence. Item validity results will be discussed in connection with homogeneous scale validation in a later section.

Construction of Homogeneous Keys

Scaling Procedure

To determine whether the items in the Crew Interaction Scale could be classified into several independent aspects of crew behavior, an analysis of the relationships among the items was made to identify homogeneous keys. The process was begun by classifying items into six categories, based on the rationale for test construction. Items were chosen for inclusion in the first tentative clusters on the basis of their agreement indexes, or significance levels of validity, or both. *The following set of symbols was used for designating scores:*

a. The subscript *t* following the item number indicates the sum of all ratings made on the item (e.g. 1_t (The number "1" refers, of course, to item-number.)

b. The subscript *t* following the abbreviation for a crew position for items 1 to 27 indicates the sum of the ratings received by that crew position (e.g., 1 AC_t).

c. The subscript *s* following the abbreviation for a crew position for items 1 to 27 indicates the self-rating by crew position.

d. The combination of the abbreviations of two crew positions separated by an oblique line indicates the rating(s) given by the one crew position to the other (e.g., 1 AC/O).

e. On items 28 to 44 the crew positions' abbreviations following the item number indicate the ratings given the crew by that crew position (e.g., 28 AC).

There were six tentative clusters identified at the beginning of the scaling procedure. They were composed of the scores listed below.

Cluster	*Scores*[2]
1. Leadership	1_t, 3 AC_t
2. Cooperation	2_t, 5_t
3. Interpersonal Relations	12 AC_t, 25 AC_t
4. Technical Competence	4 O_t, 15 O_t
5. Morale	13_t, 33_t, 36_t, 38_t
6. Formal Group Structure	21_t, 25 P/AC, 40 P

After these preliminary scales had been assembled, scaling was done by Stice and Knoell's (6) modification of the Wherry-Gaylord (8) procedure for homogeneous keying. One advantage of this procedure is that it does not require the calculation of the item intercorrelations at the beginning of the analysis.

The correlations between the total score for each scale and score on each of the 86 items considered for inclusion in the scales were determined and evaluated. A correction was applied to the correlation obtained between the item and the scale score to compensate for spuriousness due to the inclusion of the item in the scale score. Items not in a scale, that had higher correlations with a scale than items that were included, were then added; and items having a low correlation with the scale were dropped. The process was continued in order to obtain a stable set of items for each scale. Iteration produced stable sets of items for four scales. Since the *Formal Group Structure* scale and *Interpersonal Relations* scale were not stable, no further analysis was made of them.

Approximately 26 item scores were included in the scales developed in the manner described above (see Tables 5, 6, 8, and 9). Additional homogeneous keys composed of items not already included in the existing scales were then derived. Item intercorrelations were estimated and placed in matrix form, and two additional tentative clusters were obtained by inspection. Items were added to clusters and clusters were stabilized through the iteration process. Two additional scales, identified by the titles *Crew Unity* and *Crew Coordination*, were derived (see Tables 7 and 10).

[2] See footnote to Table 1 for an explanation of the symbols.

TABLE 1

CORRELATIONS BETWEEN ITEM SCORES AND TOTAL SCORES ON THE SIX HOMOGENEOUS KEYS

Score[a]	Homogeneous Keys					
	Technical Competence	Leadership	Crew Coordination	Morale	Cooperation	Crew Unity
1_t	.20	.51	.12	.45	.57	.50
1 AC_t	.07	.61	.17	.39	.36	.33
1 AC_a	.20	.31	.01	.10	.09	.20
1 O/P	.00	.48	.15	.16	.20	.33
2_t	.21	.18	.07	.47	.67	.30
3 AC_t	−.09	.52	.28	.18	.17	.17
4 O	.62	−.11	−.17	.28	.23	.23
4 AC/O	.53	−.07	−.17	.22	.15	.13
5_t	.22	.27	.07	.30	.75	.30
5 AC	.13	.29	.04	.33	.61	.17
5 O	.36	.29	.20	.39	.61	.37
6 AC/O	.36	−.18	−.17	.22	.09	.00
6 AC/P	.16	−.22	.01	.20	.20	−.07
7_t	.13	.33	.28	.39	.25	.30
8 AC	−.07	.31	.48	.33	.04	.30
8 P/AC	.04	.19	.59	.40	.11	.17
8 P_a	−.14	.16	.48	.36	.14	.36
9 P/O	.13	.27	.28	.35	.44	.23
10_t	.22	.16	.01	.37	.15	.27
10 O/P	.07	.09	−.12	.01	−.07	.03
11 AC_t	.18	.59	.29	.35	.44	.37
12_t	.33	.27	.20	.69	.47	.40
12 AC_t	.49	.13	.07	.54	.47	.37
13_t	.20	.18	.15	.35	.28	.23
14 O	.58	.07	−.01	.31	.28	.20
14 P/O	.36	.04	.12	.16	.20	.30
15_t	.09	.33	.28	.39	.25	.62
15 O/P	−.09	.29	.23	.31	.25	.37
16_t	.11	.38	.33	.54	.39	.50
16 P	.27	.22	.12	.56	.39	.53
17_t	.29	.44	.28	.59	.39	.43
17 O/P	−.04	.29	.25	.12	.33	.07
18_t	.56	.11	−.01	.37	.23	.07
18 P/O	.33	.27	.07	.26	.33	.20
19_t	.13	.27	.28	.35	.44	.23
20_t	.16	.27	.04	.33	.23	.33
20 AC_t	−.04	.36	.04	.24	.04	.17
20 AC_a	−.02	.31	−.07	.18	.15	.17
21 AC/O	.58	−.04	−.23	.18	.15	.13
21 AC/P	.44	−.04	.07	.47	.31	.23
22_t	.24	.24	.33	.56	.41	.40
22 O	.47	−.07	−.01	.24	.12	.07
22 O_a	.40	.02	.12	.14	.09	.03
23 O	.27	−.07	.07	.03	.01	.07
23 P/O	.18	−.31	−.09	.09	−.15	.00
24 AC	−.11	.59	.39	.37	.33	.27
24 P/AC	−.16	.57	.33	.22	.12	.27
26_t	.13	.16	.23	.49	.31	.40
26 O	.22	.18	.09	.41	.31	.68
27_t	.00	.33	.17	.37	.64	.37
28_t	.24	.07	.23	.55	.33	.30
29_t	.18	−.31	−.09	−.09	−.15	.00
31_t	.16	.20	.25	.45	.47	.50
32_t	.22	.09	.15	.59	.39	.43
33 P	.11	.33	.31	.24	.15	.33
34 AC	.16	.00	−.04	.31	.04	.33
35_t	.22	.36	.28	.70	.33	.40
36 P	.02	.18	.31	.20	.23	.33
37_t	.36	.40	.25	.70	.39	.47
38_t	.13	.24	.31	.47	.41	.27

TABLE 1 (*Continued*)

Score[a]	Homogeneous Keys					
	Technical Competence	Leadership	Crew Coordination	Morale	Cooperation	Crew Unity
38 O	−.04	.33	.45	.09	.23	.17
39 P	−.27	.33	.09	.09	.17	.17
40_t	.13	.38	.20	.43	.36	.75
41_t	.22	.33	.36	.63	.31	.40
42_t	.34	.27	.31	.45	.41	.20
42 O	.07	−.02	.07	.07	.15	.17
43 O	−.29	.31	.56	.16	.07	.10

[a] The following is the set of symbols for designating the scores in this table.
a. The subscript *t* following the item number indicates the sum of all ratings made on the item (e.g., 1_t).
b. The subscript *t* following the abbreviation for a crew position for items 1 to 27 indicates the sum of the ratings received by that crew position (e.g., 1 AC_t).
c. The subscript *s* following the abbreviation for a crew position for items 1 to 27 indicates the self-rating by crew position.
d. The combination of the abbreviations of two crew positions separated by an oblique line indicates the rating(s) given by the one crew position to the other (e.g., 1 AC/O).
e. On items 28 to 44 the crew positions' abbreviations following the item number indicates the ratings given the crew by that crew position (e.g.,28 AC).

Item-scale correlations were examined and further changes were made in scale content by eliminating items having high correlations with more than one scale.

These operations resulted in the stabilization of scales identified as *Technical Competence, Leadership, Crew Coordination, Morale, Cooperation,* and *Crew Unity.* The loadings of the item total and part-scores on the six scales are shown in Table 1.

The item content of the scales, loadings, agreement indexes, and validity significance levels are given in a later section entitled "Interpretation and Discussion of Homogeneous Scales."

Intercorrelations of Scales

To determine the degree of interdependence among the scales, the six scale scores for each crew were intercorrelated by the product-moment method.[3] The resulting correlations are shown in Table 2. An inspection of the coefficients indicates that the *Technical Competence*

[3] The items were scored so that the end of the scale that was considered to represent the most favorable responses was given a weight of zero, and the response at the least favorable end of the nine-point scale was given a weight of eight. Therefore low scores were considered "good" scores.

TABLE 2

INTERCORRELATIONS OF THE SIX SCALE KEYS (MATRIX R_s)
(N=30 B-47 CREWS)

Key	2	3	4	5	6
1. Technical Competence	.15	−.03	.57	.34	.52
2. Leadership		.62	.66	.68	.74
3. Crew Coordination			.63	.49	.68
4. Morale				.74	.96
5. Cooperation					.79
6. Crew Unity					

TABLE 3
MATRIX R_s^{-1}
(Inverse of Matrix R_s)

Key	1	2	3	4	5
1	2.4786	.3978	1.4564	−2.7437	.2077
2	.3978	2.4408	−.5329	−.8105	−.9336
3	1.4564	−.5329	2.7647	−2.4754	.3325
4	−2.7437	−.8105	−2.4754	5.8863	−1.6627
5	.2077	.9336	.3325	−1.6627	2.5881

scale has the lowest correlations with the other keys. *Crew Unity* was omitted from further consideration because of its high correlation with the *Morale* key. Of the two, the *Morale* key seemed more satisfactory. Scales were positively intercorrelated with the exception of a small negative coefficient between *Technical Competence* and *Crew Coordination*.

As a further test of independence among homogeneous scales, the inverse of their correlation matrix was computed. If the matrix could not be inverted it would be an indication that the scales were not linearly independent and some of the scales would be presumed to be combinations of others. The inverse (R_s^{-1}), as shown in Table 3, gave further support to the conclusion that the various scales (except "Crew Unity") represent essentially independent dimensions.

Scale Reliability

The reliability for each scale was determined by correlating odd and even item scores, based on the nine-point scale values. The Spearman-Brown correction for estimating the reliability of the full-length scales was also applied, and the corrected values are shown in Table 4. The reliabilities for the *Leadership*, *Crew Coordination*, and *Cooperation* keys are somewhat inflated since part scores from the same item appeared in both the odd and even total scores (e.g., two scores derived from Item 1 appear on the *Leadership* key).

With but one exception the corrected odd-even reliability coefficients are satisfactorily high for administrative use as separate scale measures. The exception is the *Leadership* key which, with a coefficient of .76, is satisfactory as a subscale within a battery.

Scale Validity

In order to assess the validity of the derived homogeneous keys, scale scores were correlated against the trichotomized performance criterion with results as shown in Table 4.

Since the means on the *Leadership* key for the middle and low groups were close together and since they both differed considerably from the mean for the high group, a biserial validity coefficient was also computed. The distributions for the two lower groups were combined and the resulting distribution contrasted with the distribution for the high group. A biserial coefficient was computed in the same manner for the *Crew Coordination* key. In both cases the biserial correlations obtained in this way were considerably higher than the corresponding triserials, indicating that the two scales differentiated the high group from the other two, rather than providing a uniform differentiation among the three criterion levels.

TABLE 4

RELIABILITY AND VALIDITY OF THE HOMOGENEOUS KEYS

Key	No. of Items	Corrected odd-even reliability	*M*	*SD*	Validity										
					Trichotomized Criterion Groups							Dichotomized Criterion Groups			
					High		Middle		Low		*r*	High		Low		*r*
					M	*SD*	*M*	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*	
1. Technical Competence	6	.89	48.4	24.38	34.8	9.60	43.3	15.33	50.8	20.11	.43[a]					
2. Leadership	6	.76	19.6	9.93	25.6	7.46	17.3	10.38	15.9	8.79	−.45	25.6	7.46	16.6	9.65	−.55
3. Crew Co-ordination	5	.89	11.2	5.86	14.3	6.71	8.0	3.44	11.3	5.14	−.23	14.3	6.71	9.6	14.77	−.48
4. Morale	7	.93	48.4	24.38	48.8	23.65	44.9	20.24	51.2	28.15	.05					
5. Cooperation	5	.91	39.9	20.71	44.9	19.66	40.3	22.19	34.4	18.77	−.24					

[a] Since the "good" end of the continuum was scored zero and the "poor" end was scored eight, the correlation is positive in sign when the mean of the high group is lowest in value.

Interpretation and Discussion of Homogeneous Scales

Technical Competence Key

Composition and definition. Five of the six scores composing the *Technical Competence* key shown in Table 5 represent ratings of the Observer by other members of the crew or by himself on certain of the items, as indicated. Three of the five scores represent total crew ratings of the Observer by all three crew members. The remaining two were ratings of the Observer by the Aircraft Commander. These ratings of the observer refer to aspects of his performance such as the frequency of errors, competence in performing the technical specialty, interest in performing the specialty, and consistency in performing the crew role. The crew's ratings of satisfaction with over-all crew performance (18 total) constitute the sixth valid item in the key.

Validity and interpretation. The key has a positive, linear relationship of +.43 with the criterion, with individual items ranging in significance from .10 to .20. The findings suggest that the greater the competence of the Observer in carrying out his assigned function, the more effective the crew, as ranked on the criterion. The finding that items that evaluate the competence of the Observer are valid is understandable, since a large portion of the criterion variance is considered to be due to the Observer's performance.

Leadership Key

Composition and definition. Five of the six scores that compose the *Leadership* key, shown in Table 6, are ratings of the aircraft commander. The sixth is a rating of the copilot. Items are ratings of the aircraft commander's ability to put crew members at ease, to make being a member of the crew enjoyable, and readiness to take initiative. The final item includes ratings of both the aircraft commander and copilot for the degree of acceptance or rejection of responsi-

TABLE 5

Technical Competence Key

Item and Rating Used	Item Description	Loading	Interrater Agreement (Horst)	Contingency Coefficient	Significance Level of Validity (based on χ^2)
4 O (Total)	Rank the members of your crew for the frequency with which they make errors in the performance of crew activities in the air.	.61	.50	.50	.20
14 O (Total)	Rank the members of your crew for their competence, in air activities, in the specialty to which they are assigned.	.55	.62	.57	.10
21 AC of O (Total)	Rank the members of your crew in terms of the extent to which they are interested in their technical specialty in both air and ground activity.	.55	—[a]	.44	.20
18 (Total)	Rank the members of your crew for their satisfaction with the crew's over-all performance in both air and ground activities.	.50	.81	.50	.20
4 AC of O	(See 4 above).	.50	—	.54	.10
22 O (Total)	Rank the members of your crew for the extent to which they are consistent (level of performance same from day to day) in their performance of activities in the air.	.48	.57	.45	.20

[a] Dashes in the Agreement Index column indicate that it was inappropriate to compute this value since only one rating is involved in the score. Dashes in the Validity Level column indicate that the score was not valid at the .20 level of significance.

TABLE 6
LEADERSHIP KEY

Item and Rating Used	Item Description	Loading	Interrater Agreement (Horst)	Contingency Coefficient	Significance Level of Validity (based on χ^2)
1 AC (Total)	Rank the members of your crew for the extent to which they accept or reject responsibility in both air and ground activities.	.61	.42	. .	.1C
24 AC (Total)	Rank the two other members of your crew in terms of your liking for them in both air and ground relationships.	.59	.55	. ?	.02
11 AC (Total)	Rank the members of your crew for the extent to which they put your crew at ease and make being a member of the crew more enjoyable in both air and ground activities.	.59	.35	.46	.20
24 P of AC	(See 24 AC above)	.57	—	.66	.02
3 AC (Total)	Rank the members of your crew in terms of the extent they take the initiative in getting things done in both air and ground activities.	.52	.51	.56	.10
1 O of P	(See 1 AC above)	.48	—	.64	.05

bility. The key seems to contain important aspects of the leadership function.

Validity and interpretation. All scores included in the present composition of the key have a significant relationship with the criterion. *The key had a negative relationship with the criterion,* and a triserial validity coefficient of —.45 was obtained for the total score. The mean for the high group was well differentiated from the means of the other two groups, and a biserial correlation of the high group versus a combination of the other two yielded a coefficient of —.55.

The negative relationship can be interpreted in several different ways. For the most effective crews, functional leadership may be based in the group as a whole rather than being exercised by the Aircraft Commander. Under these circumstances the Aircraft Commander would be less frequently required to put the crew at ease, to take initiative, or to accept responsibility. An alternative explanation of the negative correlation or inversion in mean ratings is that the members of the least effective crews felt threatened by a psychological test of the kind from which these data were drawn. They would feel compelled to depict their aircraft commanders as unusually competent. The least effective crews then might be less secure in admitting their own inadequacies so that they create a "halo" of effectiveness for their aircraft commander by giving him more extreme ratings. The effect on ratings would be the same in either case. Perhaps both these types of causation occur.

A third possible interpretation is that those crews are most effective in which the aircraft commander does not put crew members at ease, does not try to make being a member of the crew more enjoyable, and does not readily accept responsibility. A similar finding has been reported by Halpin (2). He found that B-29 crews' ratings of the "consideration" of their aircraft commanders were negatively related to the effectiveness of these aircraft commanders in combat as rated by their superiors. He referred to this as "the dilemma of leadership," since the more the commander pleases his crew the lower he is rated in effectiveness by his superiors. In the case of the Crew Interaction Scale *Leadership* key, the crews that rated their aircraft commanders highest for making crew membership more enjoyable, accepting responsibility, etc., were least effective in terms of the performance criterion.

Crew Coordination Key

Composition and definition. The *Crew Coordination* key shown in Table 7 consists of ratings

TABLE 7

CREW COORDINATION KEY

Item and Rating Used	Item Description	Loading	Interrater Agreement (Horst)	Contingency Coefficient	Significance Level of Validity (based on χ^2)
8 P of AC	Rank the members of your crew for the extent to which their actions are well-timed and appropriate to the actions of the other members of the crew in air activities.	.59	—	.64	.05
43 O (Total)	Rank your crew for its "cockpit confusion."	.56	—	.57	.10
8 P	(See 8 P of AC above).	.48	—	.65	.05
8 AC	(See 8 P of AC above).	.48	.11	.57	.10
38 O (Total)	Rank your crew for the extent to which it sticks together in the face of disrupting factors in both air and ground activities.	.45	—	.55	.10

of the extent to which the aircraft commander's and copilot's actions were well-timed and appropriate. Ratings by the observer of the crew's "cockpit confusion," and of the extent to which the crew sticks together in the face of disrupting factors, are included in the key. The common element among these items seems to be coordination within the crew.

Validity and interpretation. This key has an apparent curvilinear relationship with the criterion, the most effective crews having given themselves the highest, the least effective crews the next lower, and the middle crews the lowest mean ratings. A relationship with the trichotomized criterion of —.23 and a biserial correlation of —.48 were obtained.

The keyed items have validities that are significant at or beyond the 10-per-cent level, with curvilinearity apparent in the relationships of individual scores to the criterion. The curvilinear relationship might be explained by assuming that the most effective crews did not rate themselves highly coordinated because of their type of work organization, as discussed in connection with the Leadership key; whereas the least effective crews did not rate themselves highly because their coordination was inferior to that of the middle criterion group. On the other hand, since the means of the two lower groups are close together, the curvilinearity may be fortuitous, with the true relationship being one in which the upper criterion group is differentiated from the other two.

Morale Key

Composition and definition. All items in the *Morale* key shown in Table 8 are based on the combined ratings of the entire crew. Five items are crew-level items, while two others represent the sum of the individual ratings by the entire crew.

The items seem to fall into two subgroups. Ratings such as morale of crew, feeling frustrated, and enthusiasm for flying in crew assignments seem to be directly related to morale. Items such as maintaining a consistent level of efficiency, giving directions with clarity, and being a well organized team seem to be related to crew integration and to technical competence. Although items seem to refer to several aspects of crew organization, the common content of the items is designated as *Morale*.

Validity and interpretation. The total score for the key has a bimodal distribution and a curvilinear relationship with the performance criterion. None of the items has a significant degree of validity.

Cooperation Key

Composition and definition. The *Cooperation* key, shown in Table 9, refers to the extent to which members work as a cooperative unit, are ready to accept and adopt crew decisions, and feel affected because other crew members have not provided them necessary information or instruction. Three of the five scores, including

TABLE 8

MORALE KEY

Item and Rating Used	Item Description	Loading	Interrater Agreement (Horst)	Contingency Coefficient	Significance Level of Validity (based on χ^2)
35 (Total)	Rank your crew for the extent to which it is a well organized team in both air and ground activities.	.70	.48	—	—
37 (Total)	Rate of morale of your crew in both air and ground activities.	.70	.59	—	—
12 (Total)	Rank the members of your crew for the extent to which they feel frustrated by being a member of this crew.	.69	.79	—	—
41 (Total)	Rank your crew for the extent to which it is able to maintain a consistent level of flying.	.63	.33	—	—
32 (Total)	Rank your crew for its enthusiasm for flying in present crew assignments.	.59	.64	—	—
17 (Total)	Rank the members of your crew for clarity in giving directions both in air and ground activities.	.59	.63	—	—
28 (Total)	To what extent does your crew work as a well organized team rather than as a collection of individuals in both air and ground activities.	.55	.60	—	—

TABLE 9

COOPERATION KEY

Item and Rating Used	Item Description	Loading	Interrater Agreement (Horst)	Contingency Coefficient	Significance Level of Validity (based on χ^2)
5 (Total)	Rank the members of your crew for the extent to which they work cooperatively with the crew in both air and ground activities.	.75	.77	—	—
2 (Total)	Rank the members of your crew in terms of being agreeable and willing to go along with crew decisions in both air and ground activities.	.67	.69	—	—
27 (Total)	Rank the other two crew members in terms of how often you are handicapped in doing your job because they have not given you necessary information or instruction in both ground and air activities.	.64	.45	—	—
5 O (Total)	(See 5 above).	.61	.38	.50	.20
5 AC (Total)	(See 5 above).	.61	.52	.46	.20

the ratings of the crew, the aircraft commander, and the observer, are based on the cooperation item.

Validity and interpretation. The key has linear and negative relationship with the criterion. The correlation is —.24, and only three items have a validity relationship within the .20 level of significance. To the extent that a significant trend is established in these data, the more efficient crews rate themselves lowest on cooperation, agreeableness, and feeling handicapped through lack of communication. As has been indicated on other keys, the negative validities may indicate the high-performance crews are not so centrally organized as the low-performance crews. Alternatively, they may be less defensive in completing a questionnaire and, consequently, feel less compelled to rate themselves highly. In any case this key has a moderate, negative relationship with the criterion.

Crew Unity Key

Composition and definition. The items in the *Crew Unity* key, shown in Table 10, refer to effectiveness in crew planning, the extent to which members back up one another, the desire to do a good job, and the feeling of being left out of important crew decisions. The common element is identified as crew unity.

Validity and interpretation. Because of its high intercorrelations with several of the keys (e.g., +.96 with the *Morale* key) and because it contained four items, only one of which is valid at the .20 level, this key was not considered sufficiently independent to merit separate validation or other analysis.

Further Analysis

Second-Order Analysis of Scale Intercorrelations

To determine the manner in which the five homogeneous scales were grouped at the second-order level of complexity, the scale intercorrelations were subjected to factor analysis. Two centroid factors were extracted from the correlation matrix (Table 2). The centroid and rotated loadings for the two relatively independent second-order factors derived are presented in Table 11. Examination of the rotated loadings indicates that *Crew Coordination* and *Leadership* keys are relatively pure measures of one of the second-order factors. *Technical Competence* is a relatively pure measure of the other second-order factor. The *Morale*

TABLE 10

Crew Unity Key

Item and Rating Used	Item Description	Loading	Interrater Agreement (Horst)	Contingency Coefficient	Significance Level of Validity (based on χ^2)
26 O (Total)	Rank the other two members of your crew for the readiness with which you would expect them to back you up if a misunderstanding arose in your crew in ground activities.	.68	.39	.49	.20
15 (Total)	Rank the members of your crew in terms of the extent to which they want to do a good job in both air and ground activities.	.62	.56	—	—
29 (Total)	To what extent do you feel "left out" of important decisions that your crew makes in both air and ground activities?	.62	—.06	—	—
40 (Total)	Rank your crew for its effectiveness in planning together in both air and ground activities.	.75	.11	—	—

TABLE 11

SECOND-ORDER FACTOR LOADINGS OF THE FIVE KEYS

Key	Loadings				h^2
	Centroid		Rotated		
	I	II	I	II	
1. Technical Competence	.44	−.61	.04	.75	.57
2. Leadership	.77	.36	.84	.12	.72
3. Crew Coordination	.65	.41	.77	.01	.59
4. Morale	.92	−.22	.65	.68	.89
5. Cooperation	.83	.06	.73	.40	.69

and *Cooperation* keys have loadings on both of the second-order dimensions and seem to be complex at this level. While five homogeneous keys can be identified through item-scaling, results from the second-order analysis support the conclusion that there are two underlying factors operating to determine item responses. They deal primarily with judgments of the interpersonal working relations (*Crew Coordination* and *Leadership* keys) and with judgments of the technical competence of the Observer (*Technical Competence* key). A graphic representation of the relationships is shown in Fig. 1.

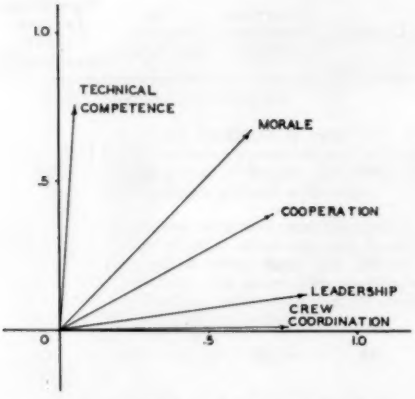

FIG. 1. Rotated positions of the five homogeneous keys on the two second-order factors.

Reliability and Validity of the Combined Scores

Since the relationships of some of the keys (*Leadership, Crew Coordination,* and *Technical Competence*) to the criterion, as well as to one another, are not linear, a combined score was obtained by an unweighted combination of scale scores, rather than by obtaining weights through multiple correlational analysis. The composite score, obtained by subtracting *Crew Coordination* and *Leadership* from the *Technical Competence* (Score = TC − CC − L), yielded a triserial validity coefficient of +.62, as shown in Table 12.

Since the high criterion group is discriminated from the other two better than they are from one another, a biserial correlation was also computed and yielded a validity coefficient of + .76. The corrected odd-even reliability of the Composite score was +.80.

Cooperation had loadings on both second-order factors, with the higher of its two loadings on the first factor. It also had a moderately negative validity coefficient. In order to determine whether the inclusion of the *Cooperation* key with its moderate validity would add to the validity of the composite, a second combination of keys was obtained by subtracting the *Cooperation* score from the composite score mentioned above (i.e., Score = TC − CC − L − C). It was anticipated that adding this key to the composite score would lower the validity of the composite, since the *Cooperation* key had loadings on factors with opposite relationships to

TABLE 12

RELIABILITY AND VALIDITY OF THE COMPOSITE SCORES

Composite Score	Corrected Odd-Even Reliability	M	SD	Validity											
				Trichotomized Criterion Groups							Dichotomized Criterion Groups				
				High		Middle		Low		r	High		Low		r
				M	SD	M	SD	M	SD		M	SD	M	SD	
TC−CC−L	.80	12.3	21.1	− 5.1	14.6	18.5	11.1	23.6	23.2	.62	− 5.1	14.6	21.0	18.4	.76
TC−CC−L−C	.92	27.5	32.2	−50.0	27.2	−21.8	25.9	−10.8	29.8	.56	−50.0	27.2	−16.3	28.5	.64

the criterion. The triserial validity coefficient for the combination of four keys was +.56, and the corresponding biserial coefficient, +.64. The corrected odd-even reliability coefficient for the combined scores from the four keys was +.92.

Based on the various combinations of scales to form a composite score, the maximum validity coefficient was given by the subtraction, without weighting, of the *Crew Coordination* and *Leadership* scores from the *Technical Competence* score.

SUMMARY AND CONCLUSIONS

The purpose of the study was to develop an improved measure of interpersonal relations among aircrew members. Based on the literature and the results from laboratory experiments designed to answer certain methodological questions, forms of the *Crew Interaction Scale* were constructed and administered to U. S. Air Force B-47 crews on two separate occasions. The initial administration, which included intensive interviews of B-47 crews to determine the suitability of items, served as the basis for the revised form. The final revised scale was administered to thirty B-47 crews of three members each and provided the data for the statistical analysis.

Agreement in Rating Items

The results of the interrater reliability analysis indicated that there is agreement among the ratings given to a crew as a whole and to its members individually, and that there is satisfactory discrimination in the average ratings for the different crews. The highest internal agreement was obtained for the first 22 items, for which nine individual judgments per item were made. The next highest internal agreement was found for items 23 through 27, that included six individual judgments per item. The least

internal agreement was found for items 28 through 44, in which only three individual judgments were represented.

Item Validity

Based on a composite performance criterion composed from rankings of crews by the Wing Standardization Board and squadron commanders, item validities were computed to provide information concerning aspects of crew relations significant for the performance criterion. Approximately 95 scores with significance levels at or below .20 for the trichotomized criterion were identified.

Homogeneous Scale Development

Scales derived from the interrelationships of the ratings lead to the identification of five dimensions of crew relations which were identified by the titles *Technical Competence, Leadership, Morale, Crew Coordination,* and *Cooperation.* The scales representing these dimensions have satisfactory reliability as determined by the odd-even method and three of them show significant validity. One of the three validly predictive scales (*Technical Competence*) has a positive relationship to the criterion. The relationship of the other two validly predictive scales (*Leadership* and *Crew Coordination*) with the criterion is *negative.*

Second-Order Analysis of Scale Intercorrelations

To determine the manner in which the five homogeneous scales were grouped at the second-order level of complexity the scale intercorrelations were subjected to factor analysis. Two centroid factors were extracted from the scale intercorrelations. The analysis of the interrelationships indicated that there are two underlying factors on which ratings are based. One can be identified with the *Technical Competence* key, which has a positive, linear, and valid relationship to the criterion. Except for further refinement, it seems to be satisfactory in its present form. The significant ratings on this scale were for the crew member called the Observer.

The other basic dimension concerns crew relationships. The *Leadership* and *Crew Coordination* keys have high loadings on it. For the *Leadership* key the high criterion group was differentiated sharply from the two lower groups. On the *Crew Coordination* key the tendency toward curvilinearity was even more marked, with the mean for the middle group being somewhat lower (representing more integration) than either extreme group.

Composite Scores

The composite scores obtained from the homogeneous keys of the *Crew Interaction Scale* take into account the opposite relationships of the basic dimensions with the criterion. A validity coefficient of .62 (triserial correlation) and .76 (biserial correlation) is obtained from the subtraction (without weighting) of *Leadership* and *Crew Coordination* from *Technical Competence.*

One other key, *Cooperation,* was considered for inclusion in a composite score since it had moderate negative validity. However, the *Cooperation* key had appreciable loading on both second-order factors, which have opposite relationships to the criterion. Combining it with the other scores raised the reliability but lowered the validity of the composite.

Conclusion

Within the framework of the present study the research has demonstrated that the direct assessment by crew members

of one another constitutes an effective means for investigating and measuring interpersonal relations in three-man crews, and provides a means for predicting the efficiency of crews that have already been organized.

An additional significant finding is that three keys, *Leadership, Crew Coordination* and *Cooperation,* were related negatively with the criterion. Possible interpretations for these unexpected results were presented in connection with the discussion of each of the keys as given in the text.

REFERENCES

1. BLAKE, R. R., MOUTON, JANE S., & FRUCHTER, B. The consistency of interpersonal behavior judgments made on the basis of short-term interaction in three-man groups. *J. abnorm. soc. Psychol.*, 1954, **49**, 415–418.
2. HALPIN, A. W. The leadership behavior and combat performance of airplane commanders. *J. abnorm. soc. Psychol.*, 1954, **49**, 19–22.
3. HOMANS, G. C. *The human group.* New York: Harcourt Brace, 1950.
4. HORST, P. A generalized expression for the reliability of measures. *Psychometrika*, 1949, **14**, 21-31.
5. LEWIN, K. *Field Theory and Social Science.* New York: Harper, 1951.
6. STICE, G. F., & KNOELL, DOROTHY M. A simple mean-difference technique for obtaining scales. *USAF Hum. Resour. Res. Cent. Res. Bulletin*, 1953, No. 53-36, iii, 53-56.
7. TAGIURI, R., BLAKE, R. R., & BRUNER, J. S. Some determinants of the perception of positive and negative feelings in others. *J. abnorm. soc. Psychol.*, 1953, **48**, 585–592.
8. WHERRY, R. J., & GAYLORD, R. H. The concept of test and item reliability in relation to factor patterns. *Psychometrika*, 1943, **8**, 247–264.

(Accepted for publication March 16, 1957)